The Emergence of Intelligent Machines: Challenges and Opportunities

CS 4700:
Foundations of Artificial Intelligence

Instructor:
Prof. Bart Selman
selman@cs.cornell.edu

The Emergence of Artificial Intelligence
Introduction

I ***Emergence of (semi-)intelligent autonomous systems in society***

--- Self-driving cars and trucks. Autonomous drones. Virtual assistants. Fully autonomous trading systems. Assistive robotics.

II ***Shift of AI research from academic to real-world***

--- Enabled by qualitative change in the field, driven in part by "Deep Learning" & Big Data.

Reasons for Dramatic Progress

--- series of events

--- main one: *machine perception* is starting to work (finally!)

systems are starting to *"hear"* and *"see"*

after "only" 50+ yrs of research…

--- dramatic change: lots of AI techniques (reasoning, search, reinforcement learning, planning, decision theoretic methods) were developed assuming perceptual inputs were "somehow" provided to the system. But, e.g., robots could not really see or hear anything…

(e.g. *2005* Stanley car drove around *blind*; developers were told "don't bother putting in a camera" --- Thrun, Stanford)

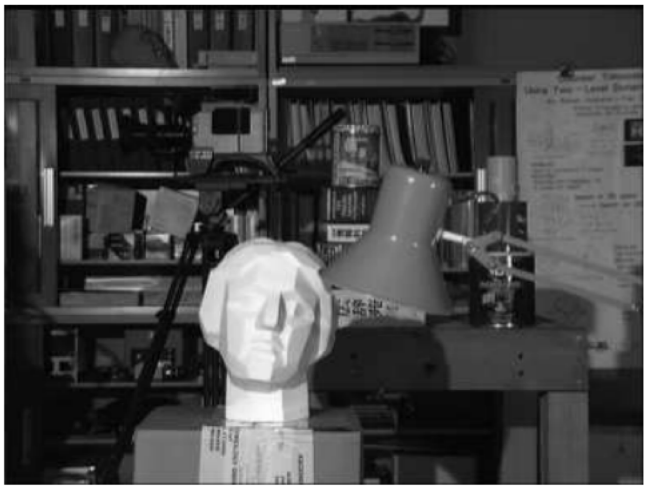
Now, we can use output from a perceptual system and leverage a broad range of existing AI techniques.

Our systems are finally becoming "grounded in (our) world."

Already: super-human face recognition (Facebook)

super-human traffic sign recognition (Nvidia)

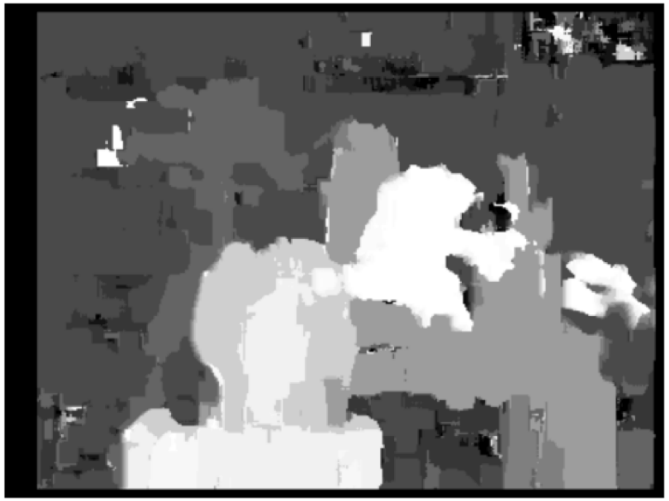
Computer vision / Image Processing ca. 2005



(a) Left image: 384x288, 15 labels

(b) Ground truth

(human labeled)

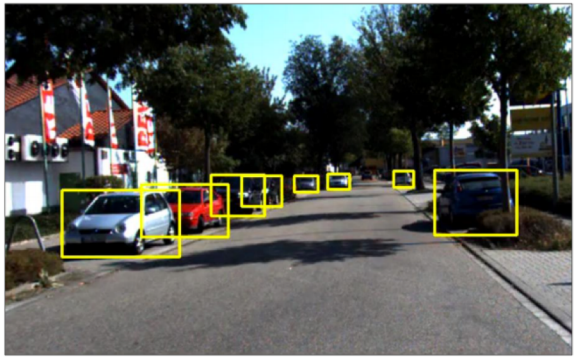


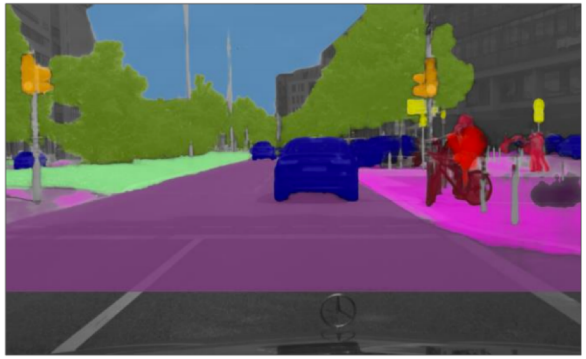
(machine labeled)

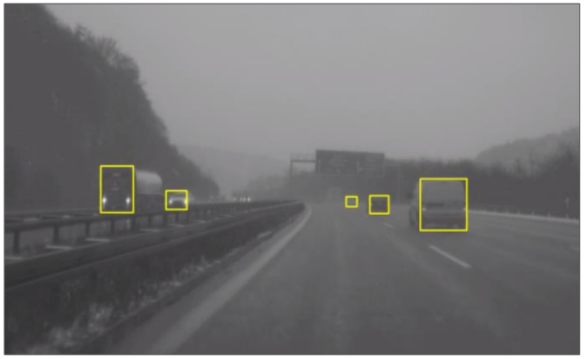
2005 --- sigh ☹

(c) Processed image

DEEP LEARNING FOR SELF-DRIVING CARS







Note labeling!

Statistical model (neural net) trained on >1M images;
Models with > 500K parameters
Requires GPU power

(Mobileye 2016; Nvidia 2016)

Real-time tracking of environment (360 degrees/ 50+m) and decision making.

Factors in accelerated progress, cont.

--- deep learning / deep neural nets

success is evidence in support of the "hardware hypothesis"

(need to get near brain compute power; Moravec)

core neural net ideas from mid 1980s

needed: several orders of magnitude increase in computational power and data

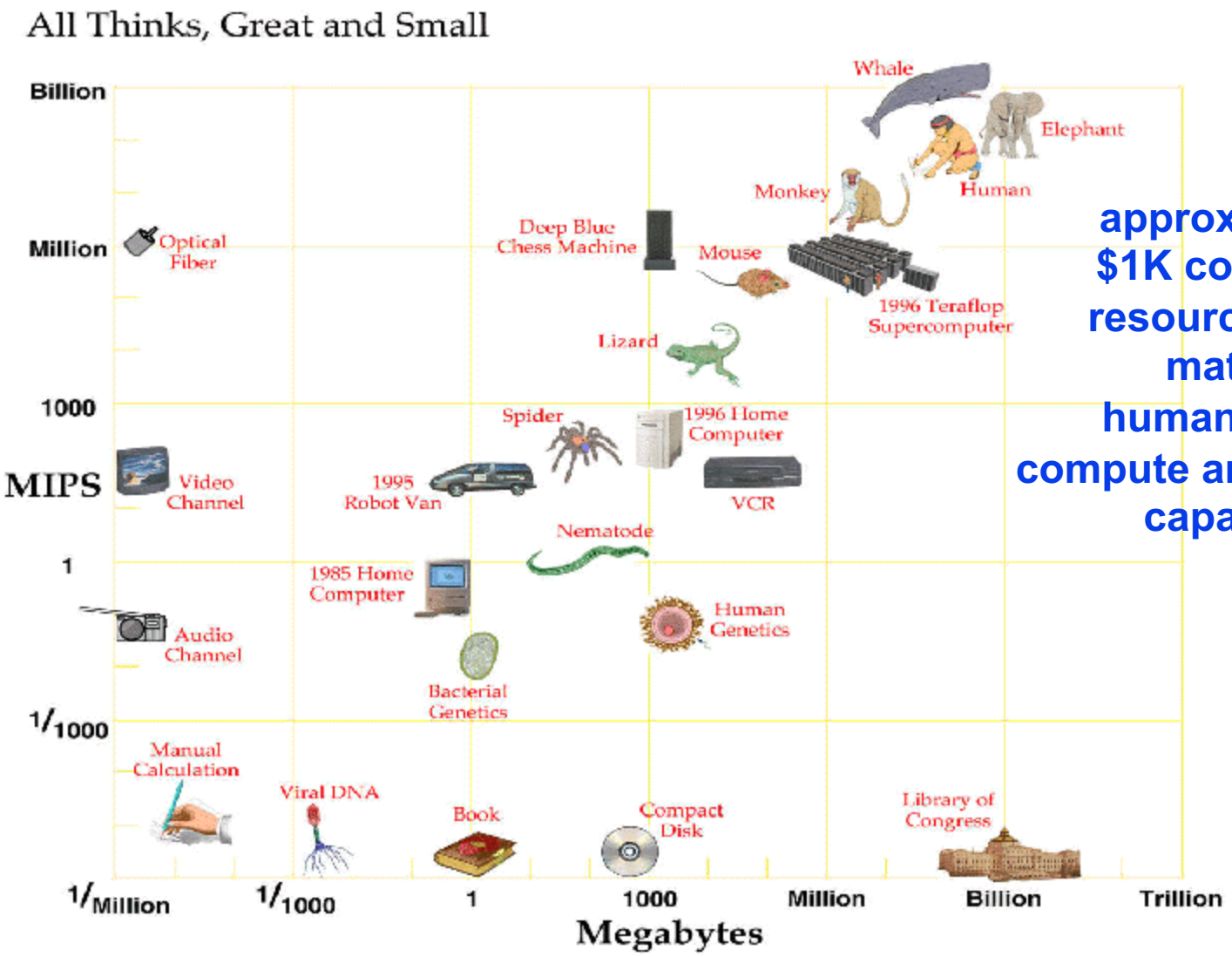
Aside:

(1) ***This advance was not anticipated/predicted at all.***
by 2000, almost all AI/ML researchers had moved away from neural nets… changed around 2011/12.

(2) Algorithmic advances still provided larger part of speedups than hardware. Core algorithmic concept from 1980s but *key additional advances since*.

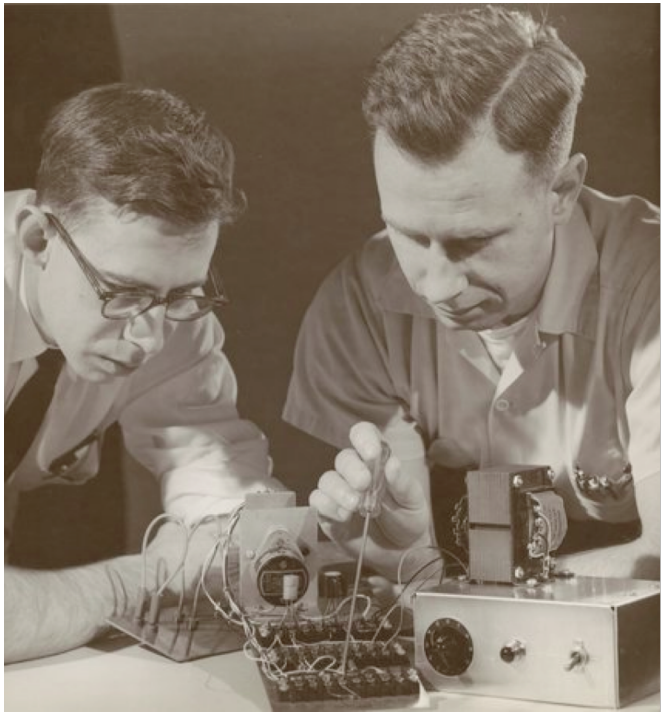
+ BIG DATA!

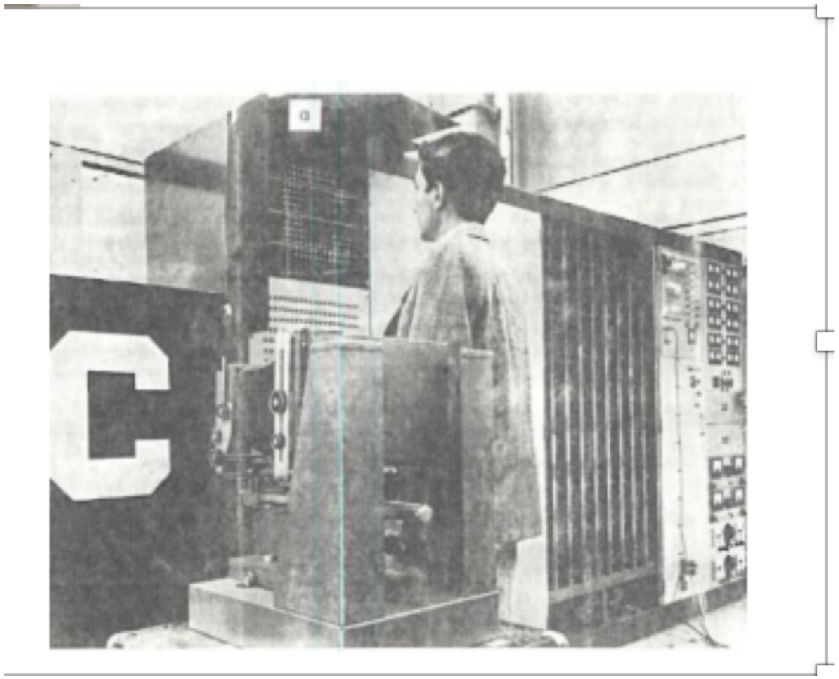
Computer vs. Brain



approx. 2030
$1K compute resources will match human brain compute and storage capacity

Historical Aside: The first learning Artificial Neural Net was developed at Cornell.





Rosenblatt (left), 1958.

(unfortunately,

patent long expired…)

Progress, cont.

--- crowd-sourced human data --- *machines need to understand our conceptualization of the world.* E.g. vision for self driving cars trained on 100,000+ images of labeled road data.

--- engineering teams (e.g. IBM's Watson)
strong commercial interests
at a scale never seen before in our field

An AI arms race

--- Investments in AI systems are being scaled-up by an order of magnitude (to billions).

Google, Facebook, Baidu, IBM, Microsoft, Tesla etc. ($2B+)

+ military ($19B proposed)

AI milestones starting in the late 90s

Lecture topic

1997 IBM's Deep Blue defeats Kasparov — Alpha-beta search

2005 Stanley --- self-driving car (controlled environment) — A* search

2011 IBM's Watson wins Jeopardy! (question answering) — K&R / agents

2012 Speech recognition via "deep learning" (Geoff Hinton) — Neural nets

2014 Computer vision is starting to work (deep learning) — Deep learning

2015 Microsoft demos real-time translation (speech to speech)

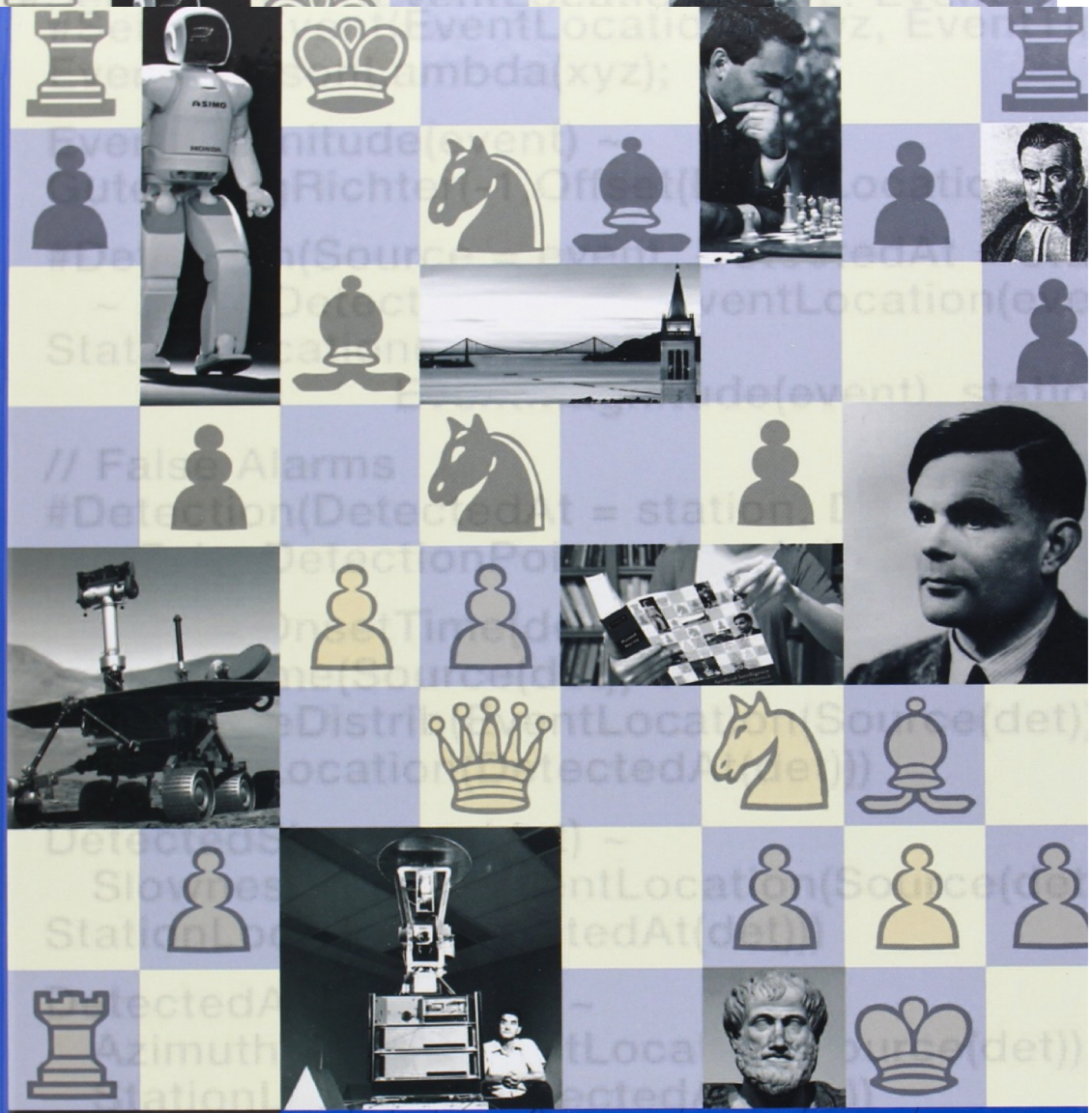
2016 Google's AlphaGo defeats Lee Sedol — Monte-Carlo search

Google's WaveNet --- human level speech synthesis — Reinforcement learning

2017 Watson technology automates 30 mid-level office insurance claim workers, Japan (IBM).

Automated dermatologists, human expert accuracy (Stanford)

Poker, Heads-up, No-Limit Texas Hold'em, CMU program beats top human players — Multi-Agent Systems



Stuart
Russell
Peter
Norvig

Artificial Intelligence
A Modern Approach
Third Edition

Historical aside:

World's first collision between fully autonomous cars (2007)

CORNELL MIT

Next Phase

Further integration of techniques --- perception, (deep) learning, inference, planning --- *will be a game changer for AI systems.*

Example: AlphaGo:

Deep Learning

+

Reasoning

(Google/Deepmind 2016, 2017)

What We Can't Do Yet [Detailed]

Aside:
Google translation is really done without any understanding of the text! (very unexpected)

Need deeper semantics of natural language

Requires commonsense knowledge and reasoning

Example:

"The large ball crashed through the table because *it* was made of Styrofoam."

What was made of Styrofoam? The large ball or the table?

"The large ball crashed through the table because *it* was made of steel."

Hmm… Can't Google figure this out? No! (Carla Gomes)

English to French (from Carla Gomes)

The vase crashed through the table because **it** was made of *steel*.
Le* vase** s'est écrasé à travers ***la table parce **qu'il** était en *acier*.

The vase crashed through **the table** because **it** was made of *Styrofoam*.
Le* vase** s'est écrasé à travers ***la table parce **qu'il** était fait de *polystyrène*.

What We Can't Do Yet

Need deeper semantics of natural language

Commonsense knowledge and reasoning

Example: "The large ball crashed right through the table because *it* was made of Styrofoam."

What does "it" refer to? The large ball or the table?

vs: "The large ball crashed right through the table because *it* was made of steel."

(Oren Etzioni, Allen AI Institute)

Commonsense is needed to deal with unforeseen cases.
(i.e., cases not in training data)

China Tesla crash --- consider how human driver handles this!

You Tube: Tesla crashes into an orange streetsweeper on Autopilot –Chinese Media

The emergence of intelligent autonomous machines among us is expected to have a major impact on society.

"Preparing for the Future of Artificial Intelligence"

White House Report,

Executive Office of the President, Oct. 2016

Societal issues:

1) **Economics (wealth inequality) & Employment**
2) **AI Safety & Ethics**
3) **Military Impact (Smart autonomous weapon systems)**
4) **The Future: Super-Intelligence? Living with smart machines.**

Elon Musk: Future of Life Institute (Max Tegmark, MIT)

AI Safety research program

In detail in "AI, Society, and Ethics" seminar course.

1) Economic Impact: Technological Unemployment

Example 1: self-driving vehicles (5 - 10 yrs).
90+% accident reduction BUT

Transportation covers about 1 in 10 US jobs!
Not so easy to replace… Also hospital emergency room reduction…

Retrain? But for what?
Knowledge worker? (see next)
STEM field? (too small)

Example 2: IBM Watson style automation of 30 insurance admin jobs (2017, Japan).

Expensive to create system but easy to duplicate…
Places *mid-level knowledge-based* jobs at risk.

Most jobs with a significant routine component will be affected.
Significant economic incentive for companies to pursue automation.
40+% of jobs at risk.

It appears inevitable that advanced AI *(systems that can hear, see, reason, plan, and learn)* will have a significant impact on employment and our society in general.

Human society will need to prepare itself.
Universal basic income?
Without work, how do we feel useful?
Amplification of wealth inequality?

2) AI Safety & Ethics

Area 1: Issues with Machine Learning (ML) Data-Driven Approaches

Data-driven ML approaches are starting to provide decision support at all levels of society.

Examples:

a) Financial loan approvals
b) Hiring / interview decisions
c) Google search order rankings
d) College applicant selection
e) Medical diagnosis
f) What's in your news feed…
g) Your year-end raise

Etc.

What about hidden biases in these decisions? & Are data-driven decisions fair?

ML approaches include hidden biases from data (e.g. past hiring / performance data) and from algorithms (e.g., what types of unfair bias cannot be eliminated?)

EU on the forefront: Working on laws to require explainable machine learning results. Also, statistical models need to be shown to adhere to non-discrimination laws.

Problem: not so easy to do!

But, at least, Google can longer just say "Results are fair because they are decided by an algorithm and data. And, algorithms and data are always fair." That worked great for a while… :-)

2) AI Safety & Ethics, cont.

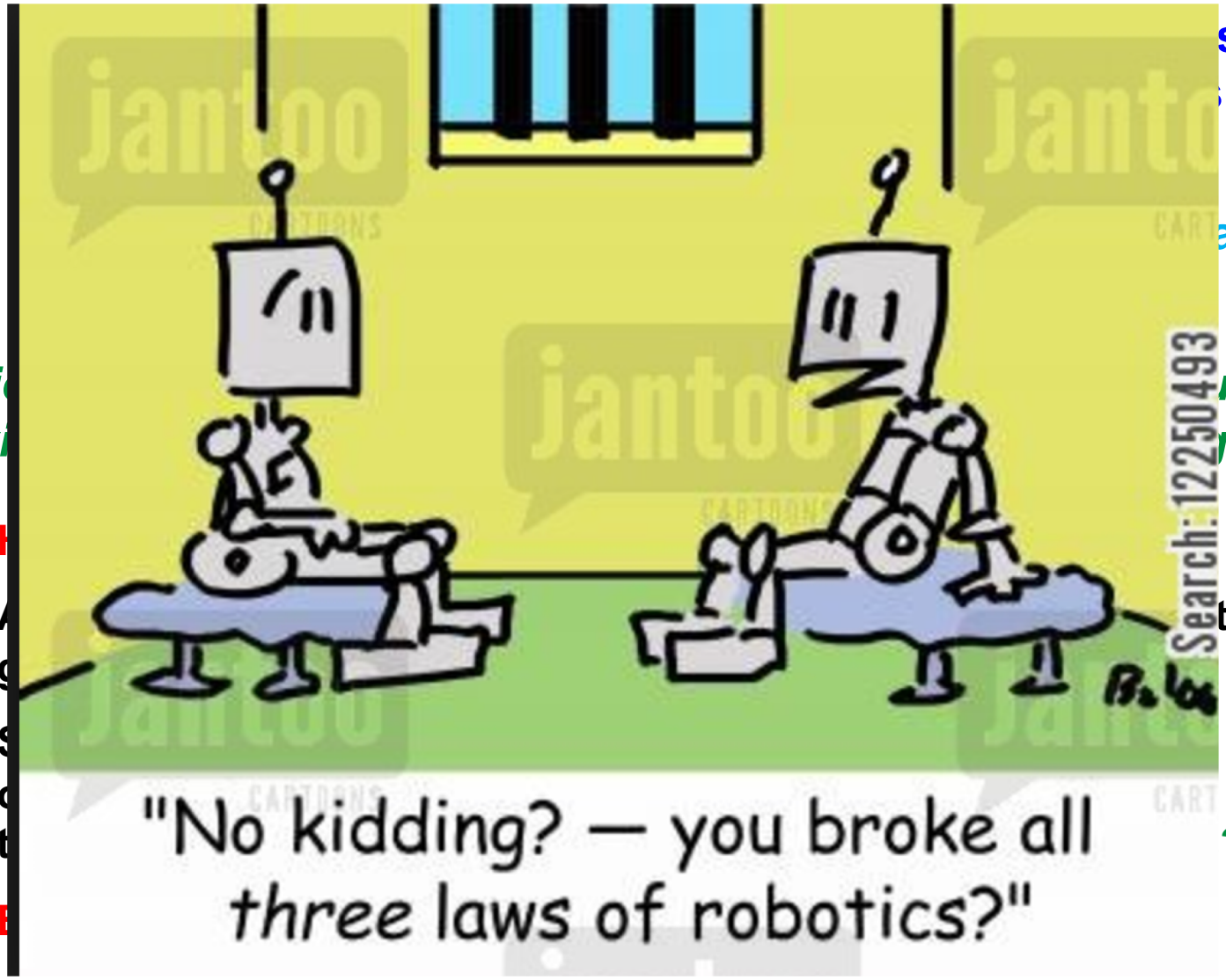
Area 2: Autonomous Goal-Driven Systems that Plan and Reason

Autonomous AI systems (eg robots or virtual assistants) no longer follow the traditional programming paradigm with detailed hand-coded sequence of instructions.
[See AI Planning in R&N.]

Instead: ***only high-level goals or instructions are given, and the system synthesizes a sequence of actions to perform given the current sensory inputs.***

How do we ensure that these decision making systems do what we want them to do and do so in a responsible matter benefiting humans?

"The Value Alignment Problem." Stuart Russell, UC Berkeley.


"No kidding? — you broke all three laws of robotics?"

3) War & Peace

AI scientists and others have recently raised significant concerns about the risks of an smart, AI-based Autonom

Lots of pre eapon systems because o ons.

Also part o ions, with countries

Issue far fr national and internatior ies are being considered. Call for Autonomous AI Weapon Ban. August 2017.

AI researchers discussing the risk of an AI Arms Race at the White House, 2016.

4) Future: Super-Human Intelligence?

Stephen Hawking: AI will be 'either best or worst thing' for humanity

Super-human AI often gets the most press.
Will we be "superseded" by smart machines?

May work out much better than some have argued. Push for AI Safety Research (funded by Elon Musk and others) will quite likely ensure a tight coupling between human and machine interests.

Also, even if machines outperform us on a range of intellectual tasks, that does necessarily mean we won't be able to understand the systems. *Humans can understand complex solutions even if we do not discover them ourselves!*

We're on an exciting intellectual journey in the history of humanity!